Functional Principle Component Analysis

Pang Du

Department of Statistics
Virginia Tech

About principle component analysis

- PCA is usually used when we want to find the dominant *modes of variation* in the data, usually after subtracting the mean from each observation.
- We want to know how many of these modes of variation are required to achieve a satisfactory approximation to the original data.
- It may be assumed that keeping only dominant modes will improve the signal-to-noise ratio of what we keep.
- We usually want to know what these modes represent in terms that we can explain to non-statisticians. Rotation of the principal components can help at this point.

Multivariate PCA: Step one

- Find *principal component weight vector* $\boldsymbol{\xi}_1 = (\xi_{11}, \ldots, \xi_{p1})'$ for which the *principal component scores*

$$f_{i1} = \sum_{j=1}^{p} \xi_{j1} x_{ij} = \boldsymbol{\xi}_1' \mathbf{x}_i$$

maximize $N^{-1} \sum_i f_{i1}^2$ subject to

$$\sum_{j=1}^{p} \xi_{j1}^2 = \|\boldsymbol{\xi}_1\|^2 = 1.$$

  - Motivation: By maximizing the mean square, we are identifying the strongest and the most important mode of variation in the variables.

Multivariate PCA: Further steps

- Next, compute weight vector $\boldsymbol{\xi}_2$ with components ξ_{j2} and principal component scores f_{i2} maximizing $N^{-1}\sum_i f_{i2}^2$, subject to the constraint $\|\boldsymbol{\xi}\|^2 = 1$ and the additional constraint

$$\sum_{j=1}^{p} \xi_{j2}\xi_{j1} = \boldsymbol{\xi}_2'\boldsymbol{\xi}_1 = 0.$$

  - Motivation: Find the next most important mode of variation that indicates something **new**.
- Continue with $\boldsymbol{\xi}_3, \ldots, \boldsymbol{\xi}_p$ as required.

Multivariate PCA: Some notes

- The *principle component weight vectors* $\boldsymbol{\xi}_m$'s are determined up to the sign.
- Can have up to p principle components, but only the first few often ends up to be important.
- Each variable is often centered before applying the PCA.
- The *principle component scores* f_{im} often carry useful information on individual observations.

Functional PCA: Step one

Multivariate PCA $\rightarrow$ functional PCA

Sums $\rightarrow$ integrals

- Find *principal component weight function* $\xi_1(t)$ for which the *principal component scores*

$$f_{i1} = \int \xi_1(t) x_i(t) dt$$

maximize $N^{-1} \sum_i f_{i1}^2$ subject to

$$\int \xi_1^2(t) dt = 1.$$

Functional PCA: Further steps

- Next, compute weight function $\xi_2(t)$ and principal component scores f_{i2} maximizing $\sum_i f_{i2}^2$, subject to the constraint $\|\xi_2\|^2 = \int \xi_2^2(t)dt = 1$ and the additional constraint

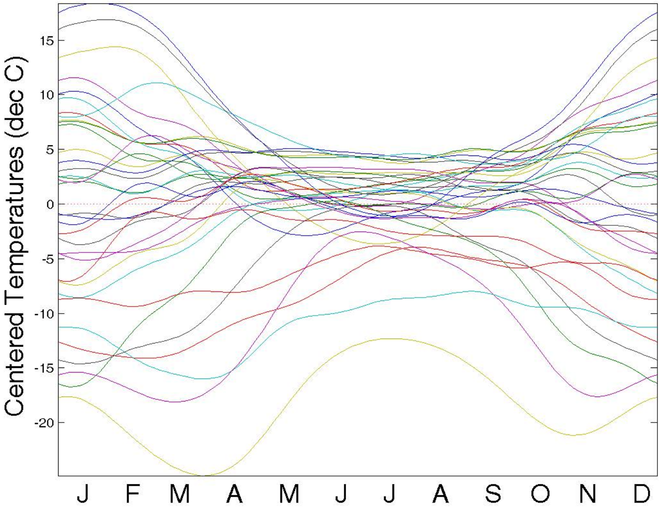
$$\int \xi_2 \xi_1 = 0.$$

- Continue with $\xi_3(t), \ldots,$ as required.

Functional PCA: An example on temperatures

- We have 30-year average temperatures for each month and for each of 35 Canadian weather stations.

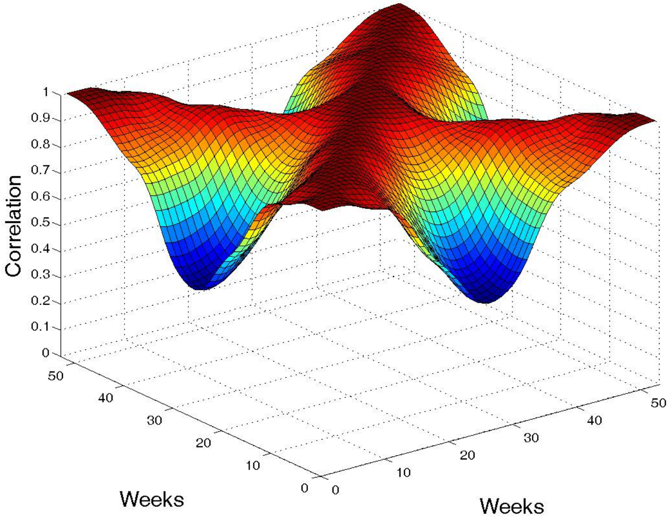
Centered monthly temperature curves



What do we see?

- An impression that some curves are high (warm) and that some curves are low (cold).
- Also that some curves have larger variation between summer and winter than others.
- How much of the variation do these two types of variation account for?

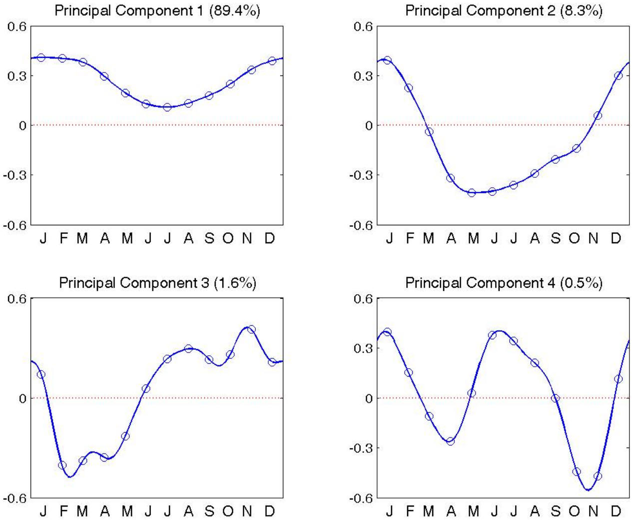
The correlation surface



What do we see?

- The diagonal ridge corresponding to unit correlation between temperatures at identical times.
- The ridge perpendicular to this corresponding to correlations between temperatures symmetrically placed around mid-summer.
- Correlations fall off much more rapidly for times symmetric about March (end of winter?) and September (end of summer?).

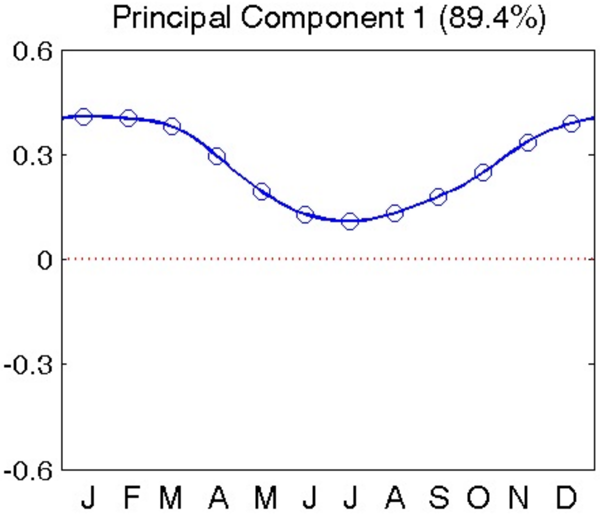
The first 4 PCs



What do we see: An overall picture

- The two components that we saw in the centered curves account for about 98% of the variation.
- The first four components account for 99.8% of the variation.
- The first four components tend to look like linear, quadratic, cubic and quartic polynomials, respectively. Why is that?
- It can help to plot the components by adding and subtracting a multiple of them from the mean function.

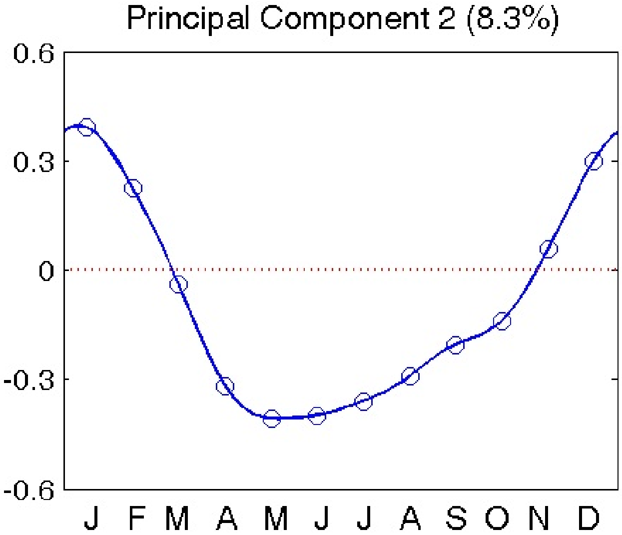
The first PC



Canadian temperature: The first PC

- The first PC is positive throughout the year, but put more weights on winter temperatures than on summer temperatures.
- Higher PC scores indicate warmer weather: highest scores among f_{i1} are assigned to Vancouver and Victoria (both on the Pacific Coast), lowest scores among f_{i1} are assigned to Resolute (in the High Arctic).
- That is, the greatest temperature variability between weather stations is found by heavily weighting winter temperatures: Canadian weather is most variable in the wintertime.

The second PC



Canadian temperature: The second PC

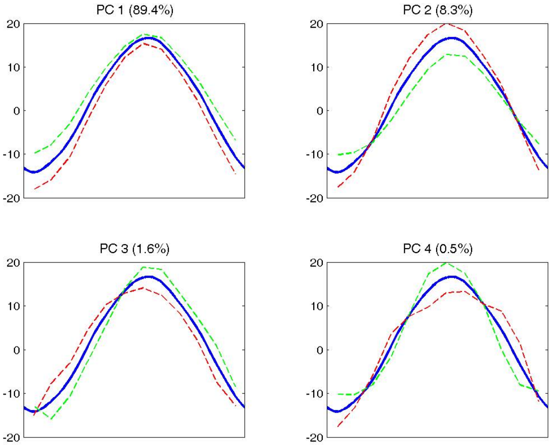
- The second PC consists of a positive contribution from the winter months and a negative contribution from the summer months.
- It corresponds to a measure of uniformity of temperature throughout the year.
- Highest scores f_{i2} go to Prince Rupert on the Pacific Coast, where there is comparatively low temperature discrepancy between winter and summer.
- Large negative scores f_{i2} go to Prairie stations (continental provinces) such as Winnipeg.

Plotting PCs as perturbations of the mean

- Obtain the estimate of the overall mean function $\hat{\mu}(t)$ of the $x_i(t)$'s.
- Create two perturbed versions of it respectively by adding and subtracting a suitable multiple of the PC function in question.
- Plot the perturbed means together with the mean estimate $\hat{\mu}$.
- Suggested perturbation constant: $0.2C$, where

$$C^2 = T^{-1}\|\hat{\mu} - T^{-1}\int \hat{\mu}(t)dt\|^2.$$

The first 4 PCs as perturbed means



Green: addition (+). Red: subtraction (−).

Canadian temperature: The 3rd and 4th PCs

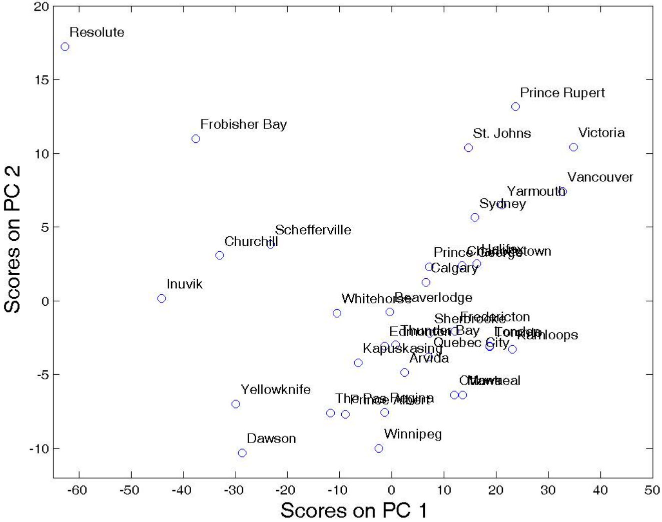
- The third PC corresponds to a time shift effect combined with an overall increase in temperature in the range between winter and summer.
- The fourth PC corresponds to an effect whereby the onset of spring is later and autumn ends earlier, as well as an increase in summer temperature.

Plotting PC scores against each other

For example, in the Canadian temperature study,

- Scores on PC1: high/low average temperatures.
- Scores on PC2: uniform/variable temperatures throughout the year.
- High scores on both PC1 and PC2: high average temperatures and uniform temperatures throughout the year (paradises like Vancouver).

PC2 vs PC1


Resolute
Prince Rupert
Frobisher Bay
St. Johns
Victoria
Vancouver
Yarmouth
Sydney
Scheffervillle
Churchill
Calgary
Inuvik
Beaverlodge
Kapuskasing
Arvida
Quebec City
Yellowknife
Dawson
Winnipeg
Scores on PC 2
Scores on PC 1
20
15
10
5
0
-5
-10
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50

What do we see?

- Most stations are along a curved line running from lower center to top right.
- At the top end of the banana are maritime stations (e.g., Vancouver) with less variation between winter and summer, and high average temperatures.
- At the lower end are the continental stations (e.g., Winnipeg) with large seasonal variation and lower average temperatures.
- The Arctic stations (e.g., Inuvik) are in their own space with large seasonal variation and very low average temperatures, except for Resolute which has consistent low temperature throughout the year.

PCs as empirical orthonormal functions

- Motivating problem: Find a set of exactly K orthonormal functions ξ_m so that the expansion of each curve in terms of these basis functions approximates the curve as closely as possible.
- Orthonormal expansion: $\hat{x}_i(t) = \sum_{k=1}^{K} f_{ik}\xi_k(t)$, where f_{ik} is the PC value $\int x_i\xi_k$.
- Fitting criterion: minimize

$$\mathrm{PCASSE} = \sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2 = \sum_{i=1}^{N} \int [x(t) - \hat{x}(t)]^2 ds.$$

- Solutions match the PC weight functions ξ_m introduced earlier, which are called *empirical orthonormal functions* here.

About empirical orthonormal functions

- Think of principal components as a set of orthogonal basis functions constructed so as to account for as much variation at each stage as possible.
- Target on approximating the data with as few basis functions as possible.
- The functions ξ_m come out looking like polynomials of increasing degree because dominant variation tends to be smooth (i.e., nearly constant or linear), and subsequent components pick up variation that declines in smoothness, and is also required to be orthogonal to previous components. Just like Legendre orthogonal polynomials!

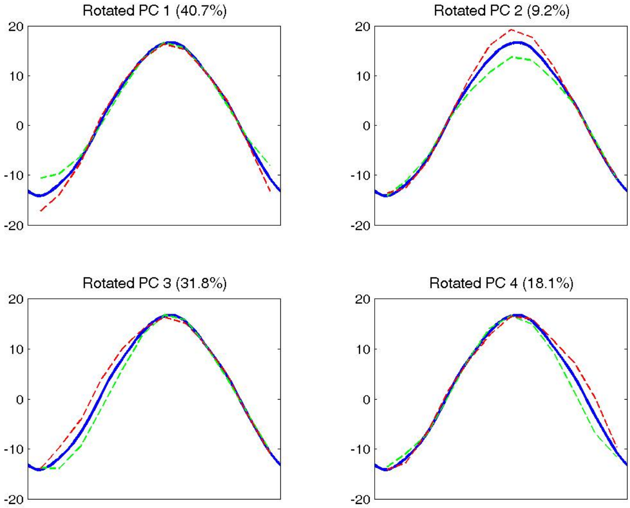
Rotating PCs: Multivariate PCA

- $\boldsymbol{\xi}_m, m = 1, \ldots, K$: orthonormal PC weight vectors explaining as much variation as we desire.
- $\mathbf{T}$: an orthonormal matrix of order K: $\mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = \mathbf{I}$.
- The rotated vectors $\mathbf{T}\boldsymbol{\xi}_m, m = 1, \ldots, K$ yield PCs that explain as much variation as those from the unrotated ones.
- VARIMAX rotation: large *loadings* (large positive/negative ξ_{mj}) for as few j as possible and many negligible loadings (ξ_{mj} close to 0) for as many j as possible.

Rotating PCs: Functional PCA

- Vector of weight functions: $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)'$.
- Rotated weight functions: $\boldsymbol{\psi} = \mathbf{T}\boldsymbol{\xi}$.
- VARIMAX rotation:
  - Take n equally-spaced points $t_1, \ldots, t_n$ such that $a_{mj} = \psi_m(t_j)$.
  - Maximize the variation within the sequence $a_{mj}^2, j = 1, \ldots, n; m = 1, \ldots, K$.
  - Note that $\sum_m \sum_j a_{mj}^2$ is fixed because $\mathbf{T}$ is orthonormal.
  - Hence encourage "sparse" a_{mj}^2 and easier interpretation.
  - Warning: the rotated function ψ_1 may NOT explain the most variation, and so on!

The rotated PCs as perturbed means



What do we see?

- The total variation accounted for remains the same, 99.8%.
- The first two components now account for a less overwhelming amount of the variation.
- Each rotated component now accounts for departure from the mean for a small part of the year.
- These are much easier to interpret: Components 1-4 account for deviations from the mean in winter, summer, spring and autumn, respectively.

Determine the number of PCs

- In the multivariate case, the upper limit is the number of variables.
- In the function case, "variables" correspond to values of t, and there is no limit to these.
- Instead, the upper limit is the number N of observations, or $N - 1$ if the functions are centered.
- But in some cases, the number of basis functions K will be less than N, and in this case K is the upper limit.
- We usually stop far short of either of these limits, however.

Computation in multivariate PCA

- we solve the eigen-equation

$$\mathbf{V}\boldsymbol{\xi} = \rho\boldsymbol{\xi},$$

  where
  - $\mathbf{V}$ is the sample variance-covariance matrix:

$$\mathbf{V} = N^{-1}\mathbf{X}'\mathbf{X}$$

    Here $\mathbf{X}$ is the centered data matrix.
  - $\boldsymbol{\xi}$ is an eigenvector of $\mathbf{V}$.
  - ρ is an eigenvalue of $\mathbf{V}$.
- Usually, however, we actually use the correlation matrix $\mathbf{R}$ instead of $\mathbf{V}$ so as to eliminate uninteresting scale differences between variables.

Eigenequation for functional PCA

- With centered functions $x_i(t)$, define

$$v(s,t) = N^{-1} \sum_{i=1}^{N} x_i(s) x_i(t).$$

- $v(s,t)$ is the sample variance-covariance function.
- The functional eigenequation is

$$\int v(s,t)\xi(t)dt = \rho\xi(s).$$

- ρ is still an eigenvalue, but now $\xi(s)$ is an eigenfunction of the variance-covariance function.
- There is much less reason for using the correlation function $r(s,t)$ since function values all have the same units or scale.

Reintroducing basis expansion

- Suppose that the observed functions are expanded in terms of a vector $\phi(t)$ of K basis functions

$$\mathbf{x}(t) = \mathbf{C}\phi(t),$$

where the coefficient matrix $\mathbf{C}$ is $N \times K$.

- and the jth eigenfunction the expansion

$$\xi_j(s) = \mathbf{b}_j'\phi(s).$$

- Substituting these expansions into the equation for $v(s,t)$ gives us

$$v(s,t) = N^{-1}\phi(s)'\mathbf{C}'\mathbf{C}\phi(t)$$

To a constrained equation

- The eigenequation becomes

$$N^{-1}\phi'(s)\mathbf{C}'\mathbf{C}\int \phi(t)\phi'(t)dt\mathbf{b}_j = \rho\phi'(s)\mathbf{b}_j$$

- Define order K matrix $\mathbf{W} = \int \phi(t)\phi'(t)dt$ so that the eigenequation is now

$$N^{-1}\phi'(s)\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b}_j = \rho\phi'(s)\mathbf{b}_j$$

- This equation has to be true for all argument values s, and consequently,

$$N^{-1}\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b}_j = \rho\mathbf{b}_j$$

subject to the constraint $\|\xi\|^2 = 1$, which becomes

$$\mathbf{b}_j'\mathbf{W}\mathbf{b}_j = 1.$$

Final solutions

- letting $\mathbf{u}_j = \mathbf{W}^{1/2}\mathbf{b}_j$ gives the eigenequation

$$N^{-1}\mathbf{W}^{1/2}\mathbf{C}'\mathbf{C}\mathbf{W}^{1/2}\mathbf{u}_j = \rho\mathbf{u}_j$$

  subject to the constraint $\mathbf{u}_j'\mathbf{u}_j = 1$.
- use standard software to solve for the eigenvectors $\mathbf{u}_j$ and back-solve to get the required coefficient vectors

$$\mathbf{b}_j = \mathbf{W}^{-1/2}\mathbf{u}_j$$

  for computing the eigenfunctions $\xi_j(s)$.

Two special cases

- Orthonormal basis $\phi_k, 1 \le k \le K$:
    - $\mathbf{W} = \mathbf{I}$,
    - FPCA reduces to standard multivariate PCA.
- Use x_i's as their **own** basis expansions:
    - $\mathbf{C} = \mathbf{I}$,
    - the problem reduces to the eigenanalysis of the symmetric matrix $N^{-1}\mathbf{W}$ with entries

$$w_{ij} = \int x_i x_j.$$

    - appropriate only when N is not large.

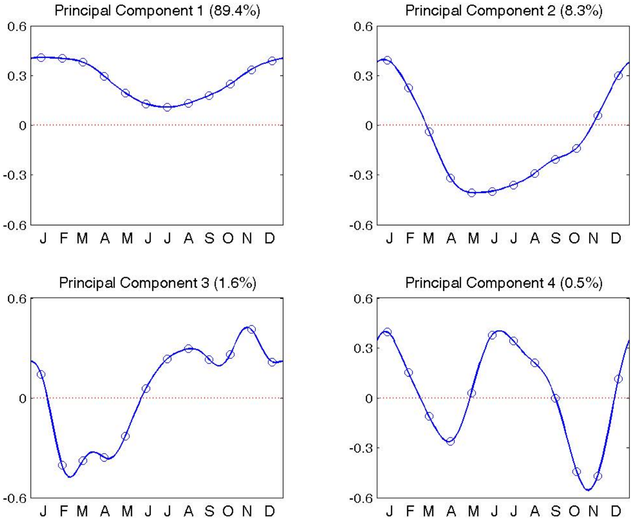
Canadian temperature example

- Temperatures were centered first:

$$x_i = \text{Temp}_i - \frac{1}{35}\sum_j \text{Temp}_j.$$

- Use $K = 12$ orthonormal Fourier series basis functions over the interval $[0, 12]$.

Canadian temperature: the first 4 PCs



Functional principal component analysis

- As simple as it sounds, the extension of multivariate PCA to functional PCA only involves a change of sums to integrals.
- Most concepts in multivariate PCA can be directly transferred to functional PCA.
- Visual tools are important and represent the FPCA results in a way much more friendly for interpretation.
- Computation can be reduced to standard matrix eigenanalysis like multivariate PCA.